

Comparison of hom(oe)ologous regions containing clusters of duplicated RGAs within *Musa* species and with rice

Stéphanie Sidibe Bocs¹, Georgios Joannis Pappas Jr², Mathieu Rouard³, Fabrice Legeai⁴, Gaëtan Droc¹, Vincent Jouffe¹, Olivier Garsmeur¹, Céline Cardi¹, Angélique d'Hont¹, Takashi Matsumoto⁵, Robert Neil Gerard Miller⁶ and Franc-Christophe Baurens¹

¹ Développement et Amélioration des Plantes, UMR1098 CIRAD, TA A 96/03, Avenue d'Agropolis, 34398 Montpellier, Cedex 5, France

{stephanie.sidibe-bocs, gaetan.droc, vincent.jouffe, olivier.garsmeur, celine.cardi, angelique.dhont, franc-christophe.baurens}@cirad.fr

² Bioinformatics Laboratory, Universidade Católica de Brasília - UCB / CAMPUS II, SGAN 916, Av. W5 Norte, CEP: 70.790-160 - Brasília/DF, Brasil
gpappas@bioinformatica.ucb.br

³ Programme Bioversity International, Commodities for Livelihood, Parc Scientifique Agropolis II, 34397 Montpellier, France
m.rouard@cgiar.org

⁴ Unité de Recherche Génomique-Info, INRA1164, centre de Versailles-Grignon, 523 place des Terrasses de l'Agora, 91034 Evry Cedex, France
fabrice.legeai@versailles.inra.fr

⁵ Rice Genome Research Program, NIAS and STAFF, 1-2 Kannondai 2-chome, Tsukuba, Ibaraki 305-8602, Japan
mat@nias.affrc.go.jp

⁶ Programa de Pós-Graduação "Stricto Sensu" em Ciências Genômicas e Biotecnologia, UCB / CAMPUS II, SGAN Qd. 916 Módulo B, Av. W5 Norte, CEP: 70.790-160 - Brasília/DF, Brasil
rmiller@pos.ucb.br

Abstract: *Understanding structure and evolution of genomic regions coding for proteins of agronomical interest is an important objective for crop improvement. We compare hom(oe)ologous regions within monocot genomes through BAC annotation. Here, we present putative orthologous and paralogous relationships of a highly duplicated Resistance Gene Analog (RGA) locus within Musa species and between Musa and rice species.*

Keywords: Comparative genomics, genome annotation, gene prediction, evolution, monocots, resistance gene analogs.

1 Introduction

Cultivated bananas are highly sterile, parthenocarpic diploid or polyploidy crops, originating from two wild diploid species: *Musa acuminata* (AA) and *Musa balbisiana* (BB). Banana cultivars are composed of clones that are vegetatively propagated and very sensitive to diseases (e.g. fungi, virus). Understanding genetic resistance to pathogens will benefit agriculture by providing a key to improving yield and reduce use of harmful products.

2 Identification of BAC Containing RGAs

BAC clones containing Resistance Gene Analog (RGA) loci using degenerated primers [1] were isolated from three BAC libraries: (i) MA4 from the wild diploid AA Calcutta 4, (ii) MBP from the diploid BB Pisang Klutuk Wulung and (iii) MAC from the most economically important triploid AAA Grande Naine (Cavendish). Analysis of BAC fingerprints and hybridizations with RGA probes allow us to select MA4_52E23 (80 kb), MBP_32N20 (141 kb) and MAC_91O16 (93 kb) which potentially contain orthologous clusters of collinear RGAs (MARGA08). BACs were sequenced by NIAS.

3 Comparative Annotation Strategy

This strategy is composed of annotation ontology, pipelines, interfaces and phylogenetics analyses, and is the fruit of previous experience with similar domains [2, 3].

Objective	Description	Reference
Gene prediction	EuGene v3.2 with rice optimized parameters v2.2	This work for parameter optimization [4]
Comparative Genomics pipeline	Integration of Blastn, Blastclust and Bidirectionnal Best Hit (BBH) methods and generation of the GenomePixelizer output format	This work [5]
Refined annotation pipeline	Integration of Blastx, Sim4, Blastp, Exonerate, InterproScan and generation of a EMBL output format	This work [6-8]
Annotation standardization	Annotation rules, decision tree, feature and qualifier definition	This work [9-12]
Annotator interfaces	Artemis, ACT, GenomePixelizer	[12, 13]
Phylogenetics analyses	Clustal, MEGA	

Table 1. Comparative annotation strategy steps.

4 Annotation of BAC Sequences Containing MARGA08 Clusters

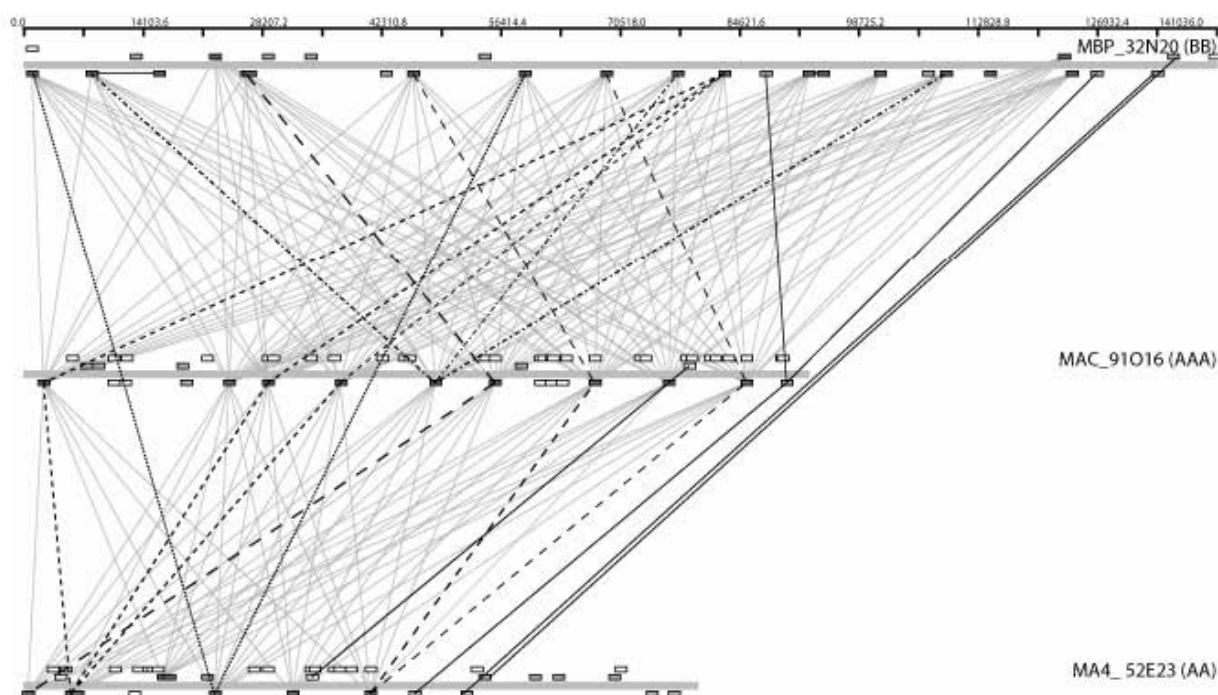


Figure 1. Putative orthologous clusters of duplicated RGAs (MARGA08) within three *Musa* clones. First Track, repeats on the BAC sequence; Second Track, genes on the direct

strand; Third Track, genes on the reverse strand; Grey Box, RGA; Pale Grey Box, non RGA; White Box, transposon gene; Dashed Line, putative orthologous relationship between RGAs; Plain black line, putative orthologous relationship between non RGAs; Plain grey line, homologous relationship between RGAs (08/04/2007).

We applied a comparative annotation strategy to these three BAC sequences:

- About 20 genes were predicted on each BAC.
- More than 50% of the largest BAC sequence, MPB_32N20, shows nucleotidic conservation (identity > 80 %) with the smaller one, MA4_52E23; a large group of about 30 homologous RGA genes was identified ($\frac{1}{4}$ MA4_52E23 genes, $\frac{1}{2}$ MBP_32N20 genes and $\frac{1}{4}$ MAC_91O16 genes); four putative clusters of orthologous RGA genes and two putative clusters of orthologous non RGA genes were found. Banana RGAs are similar to nucleotide binding site-leucine-rich repeat (NBS-LRR) disease resistance proteins which genes are grouped together on rice chromosome 11 (Os11g44960-Os11g45190).
- Structural and functional annotations of homologous gene clusters and specific genes were curated manually based on refine predictions, using genome browser and editor and following annotations rules.
- Additional comparative genomics analyses were made. Indeed, BBH method is not suitable to detect clusters of orthologous genes for large families with a fast rate of evolution. Five more consistent clusters of putative orthologous RGA genes are defined using a phylogenetics approach (neighbour joining on sequence dissimilarity). Four clusters of putative orthologous non RGA genes are defined based on genomic similarities.

5 Conclusion and Perspectives

Based upon analyses of the results obtained we can conclude that successful cloning of the orthologous MARGA08 locus was obtained in the three *Musa* species, with data suggesting that duplications and mutations leading to pseudogenization or neofunctionalization [14] occurred both before and after speciation. We are currently completing this work by identifying the other haplotypes from the three clones AA, BB, AAA using double hybridizations (MARGA08 probe and specific gene probe of the three BAC) in order to potentially compare the seven possible haplotypes of this locus (2, 2 and 3 haplotypes respectively for MA4, MBP and MAC). We will compare results with those of more stable gene-rich regions (*e.g. adh1, ers2*).

Perspectives are, firstly, to compare gene-rich regions of agronomical interest within other monocot species (*e.g. decaploid Saccharum* hybrid). Secondly, we are involved in the establishment of a eukaryotic annotation platform that satisfies: (i) manual structural, functional and comparative annotation of eukaryotic genes, (ii) database, bioanalysis pipelines and annotator Web interfaces and (iii) modularity, upgradeability, sustainability.

Acknowledgements

This work is supported by CIRAD, Bioversity, EMBRAPA, NIAS, UCB and Generation Challenge Programme.

The BAC libraries, produced through CIRAD, Bioversity, EMBRAPA and IEB projects and the research has been carried out in the framework of the Global Musa Genomics Consortium (GMGC). We also thank the Montpellier Languedoc-Roussillon GENOPOLE® for hosting the BAC library production.

Access to the Syngenta Musa EST database, donated by Syngenta to Bioversity for use within the framework of the GMGC is acknowledged.

References

- [1] R. N. G. Miller, G. J. Pappas Jr, F. C. Baurens, A. C. Ciampi, C. M. R. Santos, N. F. Martins, M. T. Souza Jr, and D. J. Bertioli. Analysis of Non-TIR NBS-LRR Resistance Gene Analogs in *Musa acuminata* Colla: Isolation, RFLP marker development, and physical mapping. In preparation, 2007.
- [2] M. Lescot, P. Piffanelli, A. Y. Ciampi, M. Ruiz, G. Blanc, J. Leebens-Mack, F. R. da Silva, Santos C.M.R, A. D'Hont, O. Garsmeur, A. D. Vilarinhos, T. Kanamori, T. Matsumoto, C. M. Ronning, F. Cheung, B. J. Haas, R. Althoff, T. Arbogast, E. Hine, G. Pappas Jr, T. Sasaki, M. T. Souza Jr, R. N. G. Miller, J.-C. Glaszmann, and C. D. Town. Molecular insights into the *Musa* genome: syntenic relationships to rice and between *Musa* species. Submitted in *The Plant Cell*, 2007.
- [3] N. Jannoo, L. Grivet, N. Chantret, O. Garsmeur, J. C. Glaszmann, P. Arruda, and A. D'Hont. Orthologous comparison in a gene-rich region among grasses reveals stability in the sugarcane polyploid genome. Accepted in *The Plant Journal*, 2007.
- [4] S. Foissac, and T. Schiex. Integrating alternative splicing detection into gene prediction. *BMC Bioinformatics*, 6, 25, 2005.
- [5] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25:3389-402, 1997.
- [6] L. Florea, G. Hartzell, Z. Zhang, G. M. Rubin, and W. Miller. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.*, 8:967-74, 1998.
- [7] G. S. Slater, and E. Birney. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6, 31, 2005.
- [8] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. InterProScan: protein domains identifier. *Nucleic Acids Res.*, 33:W116-20, 2005.
- [9] T. J. Carver, K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill. ACT: the Artemis Comparison Tool. *Bioinformatics*, 21:3422-3, 2005.
- [10] K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. Artemis: sequence visualization and annotation. *Bioinformatics*, 16:944-5, 2000.
- [11] A. Kozik, E. Kochetkova, and R. Michelson. GenomePixelizer--a visualization program for comparative genomics within and between species. *Bioinformatics*, 18:335-6, 2002.
- [12] J. D. Thompson, D. G. Higgins, and T. J. Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22:4673-80, 1994.
- [13] S. Kumar, K. Tamura, and M. Nei. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.*, 5:150-63, 2004.
- [14] R. C. Moore, and M. D. Purugganan. The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol.*, 8:122-8, 2005.